



Applying Semantic Technology to Knowledge Management

Getting Real Value out of this Cutting Edge Technology
By Christopher Bunk

WARNING

Semantic Technology is only one of many tools, techniques and technologies which can be used to implement an effective knowledge management infrastructure.

As a hot new technology its name has been used to detract attention from ineffective KM strategy, sell inappropriate software, and waste valuable IT resources on useless pilot projects.

Use only when prescribed by an experienced KM professional with a healthy regimen of sound management, good processes, and specific challenge(s) that must be addressed as the driving force.

Think to back in the day

- Imagine back to the days when data was typed and kept in files cabinets?
- What kind of challenges did they have with managing knowledge?
- Imagine how foreign of a concept being able to enter data into computers must have been.



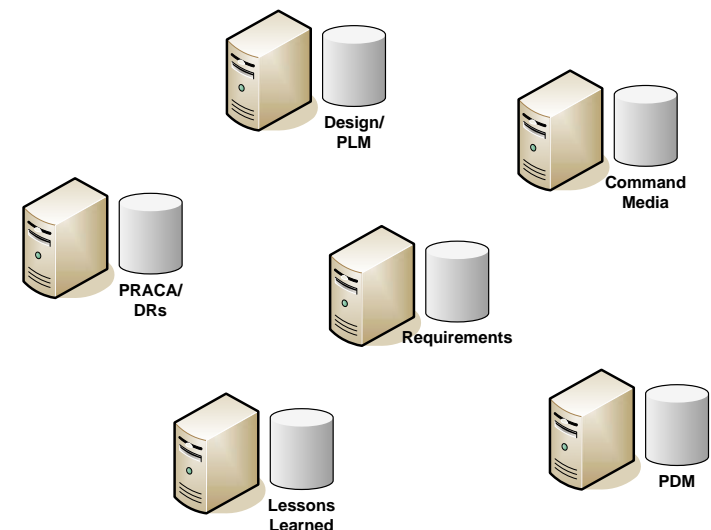
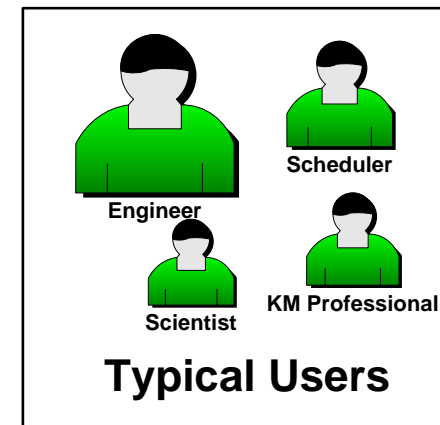
Disparate Systems

Scenario One: Using data from Disparate Systems

- *An engineer analyzes a problem report to determine a corrective action.*

Challenges:

- Collecting the data needed to solve the problem may require data from numerous systems which is very time consuming to collect and search through.
- Data in one system may not use the same naming convention as in another system.
- There is no way to do complex queries across the entire system without lots amounts of manual effort
- Searching though records containing large amounts of unstructured text requires knowing the right keywords to find what you are looking for.



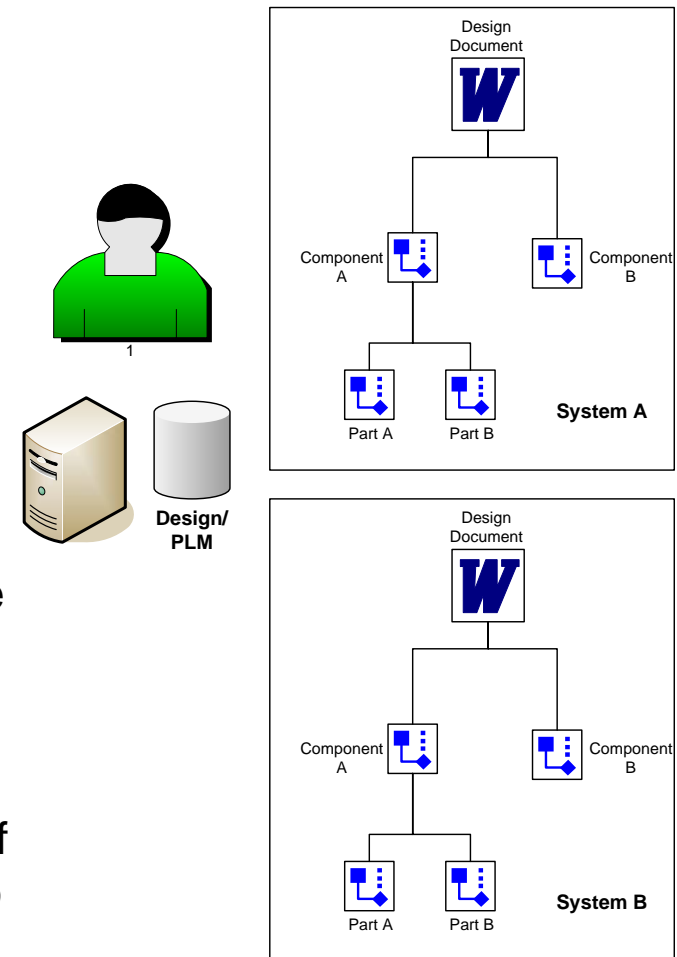
Design System

Scenario Two: Designing a new system

- An engineer has analyzed a set of requirements and is ready to sit down and crank out the design.

Challenges:

- Reusing previous designs is only possible if the engineer knows what project and part to look at.
- Searching for past Problem Reports associated with parts from past project requires searching through a different system that may use a different name or abbreviation.
- Requirements, Designs, Testing Data, Problem Reports, and Performance data are typically unlinked. If a change in a requirement is made during a late lifecycle phase it is often difficult to determine what designs need to be redone, what tests need to be performed again.
- Lessons Learned that could improve the design process typically are captured with minimal meta data. If an engineer wants to identify lessons learned relevant to what he is doing he often needs to use the right keywords relevant to what he is doing.



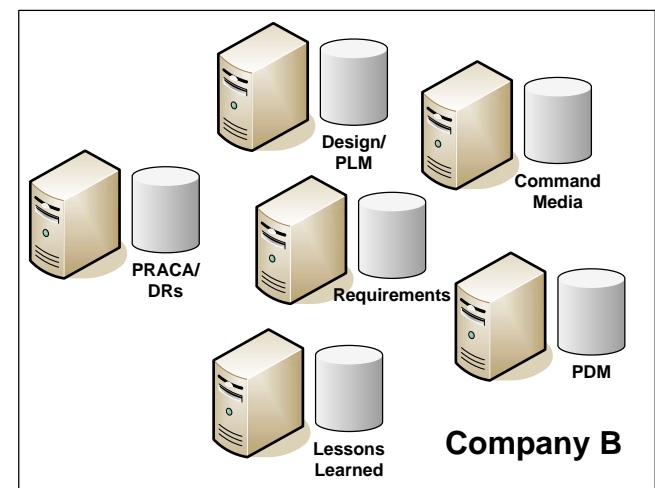
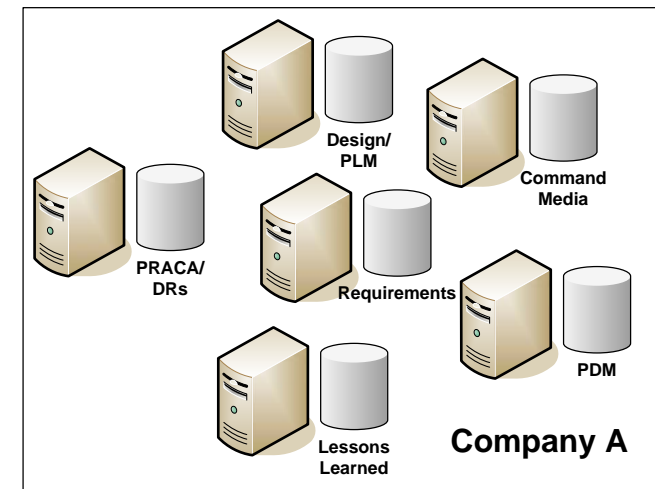
Duplicative Systems

Scenario Three: Data Integration

- Company A merges with company B. Data needs to be merged but they have different systems

Challenges:

- Manually converting all the data from one system to another requires a tremendous effort.
- Each system uses its own codes, terminology and data standards.
- Users from company B will have to learn a whole new set of tools and command media.
 - A federated solution is an option but this could be even more time consuming than converting over the data.
- Data owners can be reluctant to show their “skeletons in the closet.”



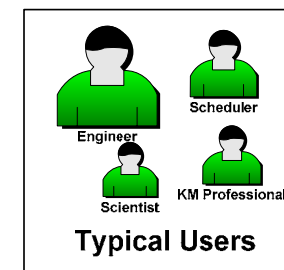
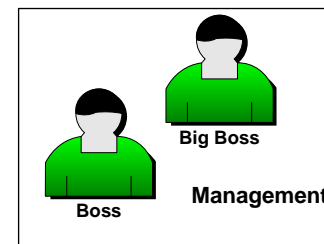
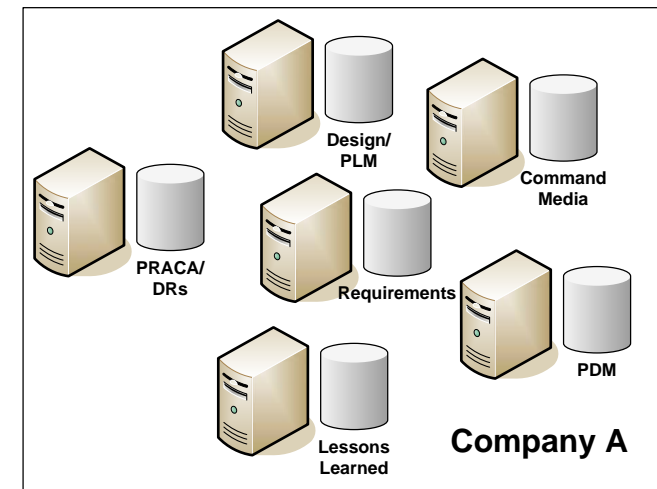
Decision Support System

Scenario Three: High Level Data Mining, Trending, Analysis and Decision Making

- *Unexpected Issue with the System gets customer upset and they want it fixed NOW!*

Challenges:

- Analysis of data across disparate systems requires an extensive effort.
- Many types of analysis not possible due to a minimal set of meta data included with data in the system.
- Some data from a subcontractor systems' may be unavailable or in a different format.
- Delivering data to a customer (including Prime Contractors and Government) on a regular basis can be labor intensive.



These problems require a new way of thinking

A corporation's data is one of its biggest assets, treat it as such.

- **Map out all of the information in your organization.**
 - Include how the data is interlinked and dependant on other data.
- **Determine what you want to be able to do with this information?**
 - Prioritize these functions according to business objectives.
 - In defining these functions you will discover additional meta data you must capture
- **Determine who will be able to see what data, and who has the ability to add, edit, and delete the data.**
- **Only once you have defined your ontology and use cases should you even begin to think about your IT infrastructure.**
 - Whether data is contained in a COTS software package, custom application, or a yet to be created system you must ensure data is accessible and represented in a way tied to your data architecture.
 - Black Box systems limit the value you realize from your data.
 - Data can be fed into a central data warehouse, be accessible through a web service as needed, or be periodically indexed.
 - Be sure to include authorization information associated with the data.

Definition of Ontology: In both computer science and information science, an ontology is a data model that represents a set of concepts within a domain and the relationships between those concepts. It is used to reason about the objects within that domain.

- Wikipedia

What do I do with my existing IT Infrastructure?

There is no need to implement a completely new infrastructure or throw away your data

- **The vast majority of COTS or custom applications store their data in a DB or have a programmatic way to get to the data such as web services.**
 - You can write programs that query this information and convert it into a format compliant with the established ontology.
- **There will be gaps that exist between your ontology and the data contained in your legacy systems**
 - There are numerous tools that exist that capable of extracting key words and the context in which they are used. Examples include a person talking about another person and command media requirement linking to another requirement.
 - If very important you could manually add meta data if cost effective.
 - There will most likely still be gaps. Use cases that require this information will not be able to be performed on this legacy data.
 - Change current systems to start collecting this additional meta data.
- **Don't forget about modifying your command media to account for changes to data systems and execution of new use cases.**
- **Train your organization on not only how to do this but why. The quality of meta data will increase if they understand why it is useful to the organization.**

**This radical change does not need to be achieved all at once.
Changes can be incrementally accomplished based on available resources and business priorities.**

What tools, techniques, and technology do I need?

Some exist, some are being developed, some have yet to be developed

- **There needs to be a way to transfer information from System A to System B**
 - Web Services over TCP/IP
- **There needs to be a way to represent this information in a tool independent format that can be understand and processed by computers**
 - Semantic Web Standards
 - XML, RDF, RDFS, OWL, SWIRL
- **There needs to be a place to store this information**
 - Triple Stores
 - Think of them as Databases of the Semantic Web
- **There needs to be a way to query this information**
 - SPARQL
 - The SQL of the Semantic Web
- **There needs to be ways to infer additional information based on the relationships and rules established in the Ontology**
 - Inference Engines
 - If Mickey is the father of Chris and George is the Father of Mickey then George is the Grandfather of Chris
- **There needs to be tools to develop the ontologies and manipulate the Semantic Data**
 - Analyze, Search, Process



Semantic Technology

The next evolutionary step of the Internet

- **In the beginning Files were linked**
 - FTP, Gopher, Archie
- **Then the WWW linked groups of text and multimedia**
 - HTML and URLs
- **The future will allow individual objects of data to be interlinked**
 - People, Places, Locations, Events, etc. can be individually represented and linked into an integrated set of data
- **Semantic Technology can be applied in a variety of ways**
 - Data Integration
 - Improved Search and Exploration of data
 - Decision Assistants
 - Business Intelligence
 - Knowledge Management
 - Content / Document Management
- **Provides the framework for data to be utilized across application and enterprise boundaries**
- **Establishes a common format for combining information from various sources of data.**
- **Data is mapped to real world objects**

"Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. Computers can adeptly parse Web pages for layout and routine processing ...but in general, computers have no reliable way to process the semantics. The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. " **- Tim Berners-Lee**



An vision of what Semantic Technology can do

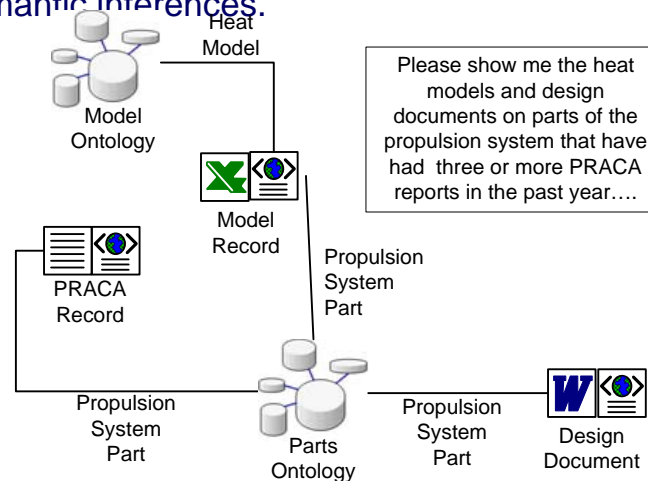
An engineering company IT infrastructure

- **Assuming that Enterprise Applications map all their data to common Ontologies you can achieve:**
 - Authorization on specific groups of data are included
 - All data in the enterprise is included such as Financial Data, Manufacturing Data, Supplier Data, News, Company Directory, Engineering Design Data, Programs, Test and Simulation, Lessons Learned, Command Media, etc.
- **Traceability of engineering data**
 - Determine which designs, test, and simulations need to be done based on a change to the requirements.
- **Rapid Prototyping**
 - Previous designs could be brought in as a template and include not the CAD drawings but the Risks, Driving Requirements, Constraints, and Lessons Learned that they are tied to.
 - Imagine the need for a jet engine with more thrust and less weight then a previous model. A previous model could be brought over and some requirements in the data could be changed. Additional data that would need to be changed for a validated model would be called out.
- **Improved reporting and exchange of data**
 - The customer could easily be provided with data deliverables in a tool independent format.
 - Automatically generated reports could pull together financial performance, supplier performance, design constraints to enable improved decision making.
- **Identify experts in specific subject areas based on the amount of documents they have authored.**
- **While working on the design of a specific part the user can use the search portal to quickly find Lessons Learned and recent news related to it.**

An vision of what Semantic Technology can do

An engineering company IT infrastructure

- Users can subscribe to be notified of new documentation related to their area that is added or updated.
 - For example a S&MA team member subscribes to a feed for any new reference documents or news that comes up related to Risk Management. When a new document is posted he will automatically be informed of the document's existence.
- **Context-aware retrieval of information**
 - Each piece of data includes a standardized XML document defining Meta attributes that describe attributes and linkages with Ontologies. In this example Meta data from the PRACA record is used to retrieve applicable models and design documents. This integration is made possible through a common parts ontology, and Semantic inferences.



Context-Aware Retrieval



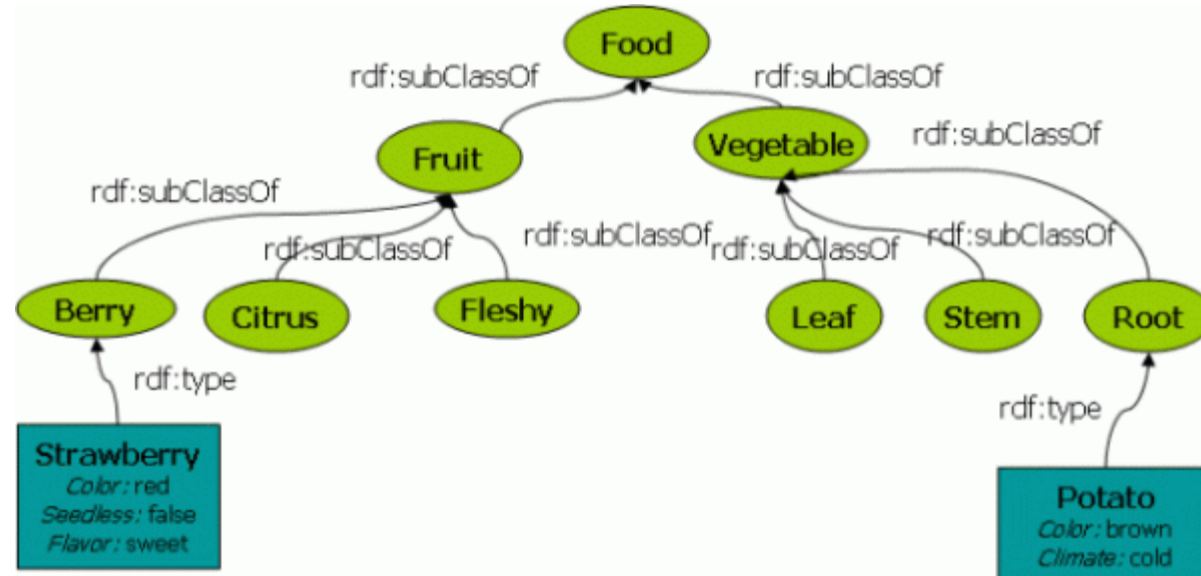
Typical Challenges and Solutions

- **Challenge:**
 - Share large volumes of data with a wide range of groups and organizations
- **Solution:**
 - Use open standards for representing data, and its associated context (meta data) through the web
 - RSS, XML, RDF, RDFS, OWL, and Web Services (SOA)
 - Use information about the data (meta data) to provide a rich, faceted approach to exploring and discovering information
 - Create Ontologies that map the inter-relationships between different types of data
- **Challenge:**
 - Collaboratively interpret and analyze engineering data
- **Solution:**
 - Create a web based system that allows workgroups and communities of practice to collaborate across geographic boundaries
- **Challenge:**
 - Integrate data from outside sources into the base of knowledge of a specific organization
- **Solution:**
 - Establish a standards based universal model (Ontology) through which information can be represented regardless of source.
 - Collect outside information through a combination of website crawling, RSS, web service data transfers, and manual uploading and convert to match the standard model.
 - Data can be explored through a semantic portal.

What is an Ontology?

As relevant to Semantic Technology

- Defines the terms used to describe and represent an area of knowledge.
- Ontologies are used by people, databases, and applications to share domain information
 - A domain is a specific subject area or area of knowledge, like Risk Management, Manufacturing, Electrical Engineering, etc.
- Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them
- A way to Infer additional relationships based on a set of rules.
 - Causal
 - Temporal
 - Probabilistic



Food Ontology from Carnegie Mellon University

The Languages that make up Semantic Technology

A set of W3C governed standards

- **URIs**
 - Global Naming Convention
 - Interpreted consistently across different context
 - Both for objects and relationships
- **RDF/RDFS**
 - Used to assign URIs to its fields
 - Expressed as Triples (Subject Predicate, Object)
 - Often Represented using RDF/XML but other representations like N3 Exist
- **OWL**
 - Greater Expression in object and relation descriptions
 - Can be used to link different ontologies
 - Used to establish rules in a domain of knowledge that can later be expanded
 - "Ontologies are attempts to more carefully define parts of the data world and to allow interactions between data held in different formats." - Nigel Shadbolt
- **SPARQL**
 - The SQL of the Semantic Web
 - Language through which programmers can interface with a triple store

Semantic Data (Triples) are stored in a Triple Store

The Database of the Semantic World

- Triples are represented using a graph model (Subject, Predicate, Object).
- Applications can query the triple store using SPARQL or a proprietary query language similar to how an application queries a database.
- Some triple stores are capable of inferencing additional triples
 - If John is son of Tom and Tom is a son of Greg and the Ontology states that a son of a son is a grandchild then the triple store may infer that John is a Grandchild of Greg without that triple being explicitly established.
- Some triple stores include
 - **Oracle**: Oracle Spatial 10g includes an open, scalable, secure and reliable RDF management platform. Based on a graph data model, RDF triples are persisted, indexed and queried, similar to other object-relational data types.
 - **AllegroGraph** (Franz): AllegroGraph is a system to load, store and query RDF data. It includes a SPARQL interface and RDFS reasoning. It has a Java and a Prolog interface.
 - **Sesame**: Sesame is an open source RDF database with support for RDF Schema inferencing and querying. It offers a large scale of tools to developers to leverage the power of RDF and RDF Schema
 - **JENA**: Jena Java RDF API and toolkit is a Java framework to construct Semantic Web Applications. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine. It also has the ability to be used as an RDF database via its Joseki layer.
 - **Seamark (Siderean)**: Seamark's triple store comes bundled with the rest of the product and uses a proprietary query language.
 - **RDF Gateway**: RDF Gateway is an RDF Triple database with RDFS reasoning and SPARQL interface
 - **Tucana** (NGC): Northrop Grumman's Tucana Suite utilizes the Kowari Metastore in its application.



Government Adoption of Semantic Technology

Through the Federal Enterprise Architecture (FEA) the government has made a strong commitment to adoption of Semantic Technology

- Used for interoperability, integration, capability reuse, accountability and policy governance in agencies, across agencies and even across governments
- <http://www.whitehouse.gov/omb/egov/a-1-fea.html>
- Department of Defense Architectural Framework (DoDAF)
 - Work underway to incorporate Semantic Languages
- NASA Constellation Data Architecture (NExIOM)
 - First appeared on CEV and CLV Programs
- As the government gets a better grasp of Semantic Technology they will require suppliers to share rich structured data.

The DARPA Agent Markup Language (DAML) program is pleased to endorse the OWL Web Ontology Language produced by the W3C Web Ontology Working Group based on the DAML+OIL language developed by the DAML program and its European Union collaborators. We view OWL as a major advancement for the Semantic Web, and have been using it extensively as part of our on-going work to develop Semantic Web tools, rules, and services. We look forward to the wide scale deployment of OWL on the World Wide Web.

- Mark Greaves, Program Manager, Defense Advanced Research Projects Agency, U.S. Department of Defense

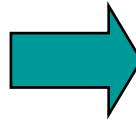
Making the first steps in adopting Semantic Technology

The standards, tools, and techniques have finally reached a mature state for the enterprise.

Pilot Project

Demonstrate the value of Semantic Technology to setup a larger success

- **Concept Based Search and Discovery**
There are mature COTS products that can be combined with custom developed Ontologies and interfaces to provide a quick benefit to the user. A good example of this would be providing an improved method of navigating engineering publications. In the cases of legacy repositories lacking the needed meta data COTS tools can be used to extract key entities from unstructured text.
- **Semantic Integration**
The value of Semantic Technology increases as the amount of data sources increase. The big win would ideally involve integration of data across the entire engineering community but a benefit can be still be achieved through integration of a few targeted legacy data sources, preferably controlled by one stake holder.
- **Science Intelligence Portal**
A specific case where an organization needs to make regular decisions with data from different sources can easily turn into a big win for the organization. An example would be a system that allows scientists to enter in criteria of a hypothesis and automatically pull in relevant data to test the hypothesis using various visualization methods.



The Big Win

By adopting semantic technology throughout an organization, return on investment can be maximized.

- **Concept Based Search and Discovery**
Similar to the pilot project but with data pulled in from across the organization.
- **Community Wide Semantic Integration**
Knowledge is the basis of engineering discovery and validation. Semantic Integration will allow anyone to utilize any combination of data from disparate sources in the most efficient manner possible.
- **Ontology Mapping**
Ontologies specific to an organization or branch of science takes time to develop and evolve. Organizations that invest in this early on will see a significant competitive advantage over other groups.
- **Engineering / Business Intelligence Portal**
Allow management, scientists, engineers and other groups have an efficient way of accessing and analyzing information from across the organization to make more effective decisions.
- **Streamlined Engineering Innovation**
Related data can be better explored, discovered, interpreted and analyzed analyzed,. Semantic Tech based knowledge tools capture, represent, and maintain total product knowledge in a language-neutral, federated repository



For more information

Let us know how we can help your group

Christopher Bunk

ARES Corporation

Knowledge Systems Engineer

o: 310-532-5677 x15

c: 440-263-2529

cbunk@arescorporation.com